



扫码试学



扫码加群

Kafka 并不难学!

入门、进阶、商业实战

邓杰◎编著



- 1 免费提供实例素材、源文件
- 2 开发一个监控系统，提供代码
- 3 短段 + 短句，便于流畅阅读
- 4 实例的商业性、应用性较强
- 5 代码均有详细注释，便于理解

图书（通过68个实例进行讲解）

+视频（115分钟，与图书同步）

+QQ交流群（作者答疑、交流）



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

Kafka 并不难学!

入门、进阶、商业实战

邓杰◎编著



电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书基于Kafka 0.10.2.0以上版本,采用“理论+实践”的形式编写。全书共68个实例。

全书共分为4篇:第1篇,介绍了消息队列和Kafka、安装与配置Kafka环境;第2篇,介绍了Kafka的基础操作、生产者和消费者、存储及管理数据;第3篇,介绍了更高级的Kafka知识及应用,包括安全机制、连接器、流处理、监控与测试;第4篇,是对前面知识的综合及实际应用,包括ELK套件整合实战、Spark实时计算引擎整合实战、Kafka Eagle监控系统设计与实现实战。

本书的每章都配有同步教学视频(共计155分钟)。视频和图书具有相同的结构,能帮助读者快速而全面地了解每章的内容。本书还免费提供所有案例的源代码。这些代码不仅能方便读者学习,也能为以后的工作提供便利。

本书结构清晰、案例丰富、通俗易懂、实用性强。特别适合Kafka系统的初学者和进阶读者作为自学教程。另外,本书也适合社会培训学校作为培训教材,还适合大中专院校的相关专业作为教学参考书。

图书在版编目(CIP)数据

Kafka并不难学!:入门、进阶、商业实战 / 邓杰编著. —北京:电子工业出版社, 2018.11
ISBN 978-7-121-35247-8

I. ①K… II. ①邓… III. ①分布式操作系统 IV. ①TP316.4

中国版本图书馆CIP数据核字(2018)第239574号

策划编辑:吴宏伟

责任编辑:牛 勇

印 刷:北京天宇星印刷厂

装 订:北京天宇星印刷厂

出版发行:电子工业出版社

北京市海淀区万寿路173信箱 邮编:100036

开 本:787×980 1/16 印张:23 字数:513千字 彩插:2

版 次:2018年11月第1版

印 次:2018年11月第1次印刷

定 价:89.00元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010) 88254888, 88258888。

质量投诉请发邮件至zltz@phei.com.cn, 盗版侵权举报请发邮件至dbqq@phei.com.cn。

本书咨询联系方式:010-51260888-819 faq@phei.com.cn。

前言

关注以下公众号，回复“KAFKA”，可获取教学视频、实例素材、实例源文件。



开源软件 Kafka 的应用越来越广泛。Kafka 简单易学，其学习曲线平缓且学习周期短。只需要较短的时间学习，就可以学会 Kafka 应用开发，完成一个高可用集群的部署和高可用应用程序的编写。

面对 Kafka 的普及和学习热潮，笔者愿意分享自己多年的开发经验，带领读者比较轻松地掌握 Kafka 的相关知识。这便是笔者编写本书的原因。

本书使用通俗易懂的语言进行讲解，从基础操作到集群管理，再到 Kafka 底层设计等内容均有涉及。本书具有以下特点。

1. 免费提供 155 分钟教学视频

作者按照图书的内容和结构，录制了同步对应的教学视频。既有上课式讲解，又有具体的代码操作。

2. 可加入本书 QQ 学习群提问、交流

本书 QQ 学习群：825943084。加入本群可与千人成为同学，共享资源。图 1 所示为群内交流情况。

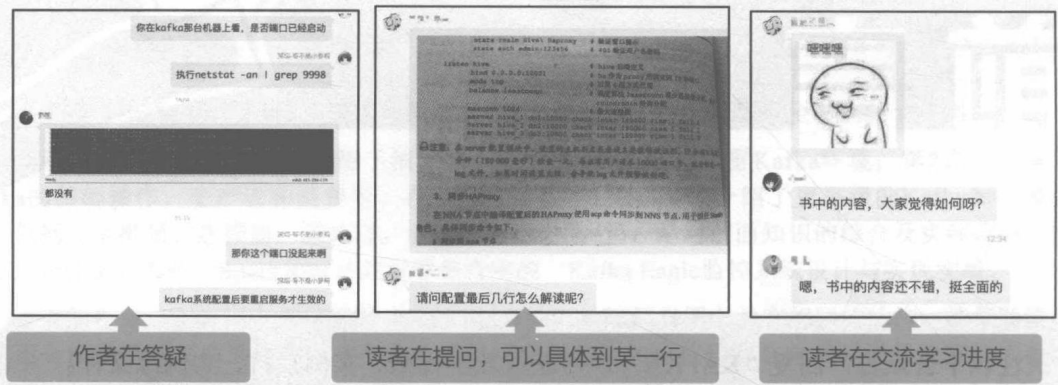


图 1 群内交流情况

3. 通过 68 个实例进行讲解

本书提供了 68 个实例，将理论讲解最终都落实到代码实现上来。而且，这些实例会伴随着图书内容的推进，不断地趋近于工程化的项目风格，具有很高的应用价值。

4. 免费提供实例素材

书中实例用到的素材已经提供，如图 2 所示。读者可以采用这些素材完全再现书中的实例效果。

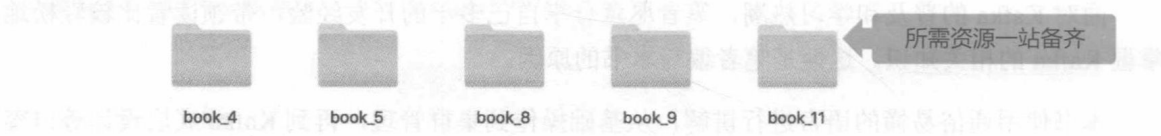
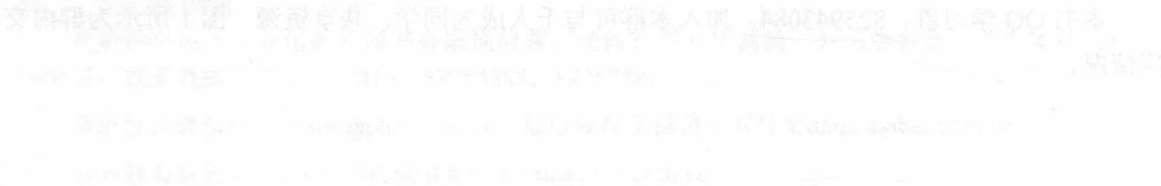


图 2 本书实例用到的素材

5. 免费提供实例的源文件

在网上已经提供了书中实例的源文件，如图 3 所示。读者可以一边阅读本书，一边参照源文件动手练习。这样不仅能提高学习的效率，还能对书中的知识有更加直观的认识，从而逐渐培养自己的编码能力。



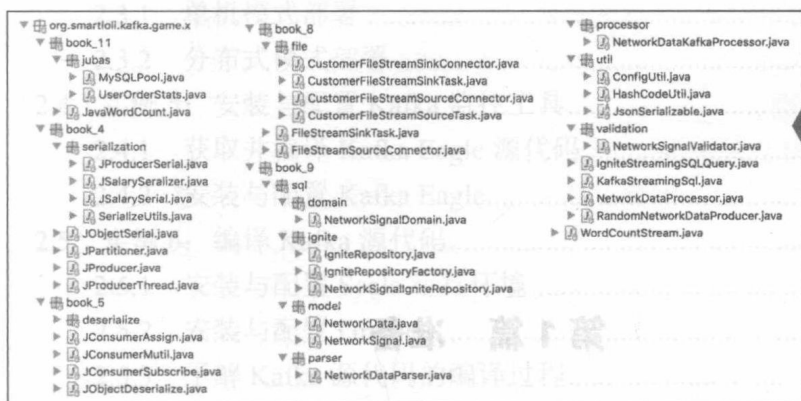


图3 本书实例源文件

6. 覆盖的知识面广

本书几乎囊括了 Kafka 0.10.2.0 版本中的全部功能。读者在系统学习之后，本书还有查阅的价值。读者可以把本书当成一本 Kafka 工具书长期保留在身边。无论何时何地，只要遇到生僻操作，都可以及时找到说明。

7. 采用短段、短句，便于流畅阅读

本书采用丰富的层次，并采用短小的段落和语句，所以，读来有顺水行舟的轻快感。

8. 实例的商业性、应用性强

本书提供的实例多数来源于真正的商业项目，具有高度的参考价值。有些代码甚至可以直接移植到实际的项目中，进行重复使用。使得“从学到用”这个过程变得更加直接。

致谢

感谢我的女朋友邹苗苗对我生活上的细心照顾与琐事上的宽容，使得我能安心写作！感谢我的父母对我的养育之恩！

感谢各位读者选择了本书！希望本书能对您的学习有所助益。

虽然我们对书中所述内容都尽量核实，并多次进行文字校对，但因时间紧张，加之水平所限，书中难免有疏漏和错误，敬请广者批评指正。

联系编辑请发 E-mail 到 wuhongwei@phei.com.cn。

目 录

第 1 篇 准备

第 1 章 了解消息队列和 Kafka	2
1.1 本章教学视频说明	2
1.2 消息队列	2
1.2.1 什么是消息队列	3
1.2.2 消息队列主要有哪些作用	3
1.3 为什么需要 Kafka	6
1.4 Kafka 的基本概念	7
1.4.1 代理、生产者、消费者、消费者组	7
1.4.2 主题、分区、副本、记录	8
1.5 了解 Kafka 的工作机制——生产消息/消费消息	9
1.6 Kafka 的使用范围	10
1.6.1 Kafka 的设计初衷	10
1.6.2 Kafka 的特性	11
1.6.3 Kafka 适用于哪些场景	13
1.7 小结	14
第 2 章 安装及配置 Kafka	15
2.1 本章教学视频说明	15
2.2 安装与配置基础环境	16
2.2.1 安装并配置 Linux 操作系统	16
2.2.2 实例 1: 安装与配置 Java 运行环境	18
2.2.3 实例 2: 配置 SSH 免密码登录	21
2.2.4 实例 3: 安装与配置 Zookeeper	23
2.3 实例 4: 部署 Kafka	27

2.3.1	单机模式部署	27
2.3.2	分布式模式部署	29
2.4	实例 5: 安装与配置 Kafka 监控工具	32
2.4.1	获取并编译 Kafka Eagle 源代码	32
2.4.2	安装与配置 Kafka Eagle	33
2.5	实例 6: 编译 Kafka 源代码	37
2.5.1	安装与配置 Scala 运行环境	38
2.5.2	安装与配置 Gradle	39
2.5.3	了解 Kafka 源代码的编译过程	40
2.6	实例 7: 将 Kafka 源代码导入编辑器	42
2.6.1	导入 IntelliJ IDEA 编辑器	42
2.6.2	导入 Eclipse 编辑器	44
2.7	了解元数据的存储分布	46
2.8	了解控制器的选举流程	48
2.8.1	了解控制器的启动顺序	48
2.8.2	了解主题分区 Leader 节点的选举过程	52
2.8.3	了解注册分区和副本状态机	59
2.8.4	了解分区自动均衡和分区重新分配	61
2.9	小结	66

第 2 篇 入门

第 3 章	Kafka 的基本操作	68
3.1	本章教学视频说明	68
3.2	操作 Zookeeper 集群	68
3.2.1	Zookeeper 的作用及背景	69
3.2.2	实例 8: 单机模式启动 Zookeeper 系统	70
3.2.3	实例 9: 单机模式关闭 Zookeeper 系统	72
3.2.4	实例 10: 分布式模式启动 Zookeeper 集群	74
3.2.5	实例 11: 分布式模式关闭 Zookeeper 集群	77
3.3	操作 Kafka 集群	77
3.3.1	实例 12: 单机模式启动 Kafka 系统	78
3.3.2	实例 13: 单机模式关闭 Kafka 系统	79

3.3.3	实例 14: 分布式模式启动 Kafka 集群	81
3.3.4	实例 15: 分布式模式关闭 Kafka 集群	84
3.4	管理主题	85
3.4.1	什么是主题	86
3.4.2	实例 16: 创建主题	87
3.4.3	实例 17: 查看主题	88
3.4.4	实例 18: 修改主题	92
3.4.5	实例 19: 删除主题	94
3.5	管理分区与副本	95
3.5.1	分区和副本的背景及作用	95
3.5.2	实例 20: 修改分区	96
3.5.3	实例 21: 修改副本数	97
3.6	小结	99
第 4 章	将消息数据写入 Kafka 系统——生产	100
4.1	本章教学视频说明	100
4.2	了解 Kafka 生产者	101
4.3	使用脚本操作生产者	101
4.3.1	实例 22: 通过监控工具查看消息	102
4.3.2	实例 23: 启动消费者程序, 并查看消息	103
4.4	发送消息到 Kafka 主题	104
4.4.1	了解异步模式	104
4.4.2	实例 24: 生产者用异步模式发送消息	105
4.4.3	了解同步模式	105
4.4.4	实例 25: 生产者用同步模式发送消息	106
4.4.5	多线程发送消息	107
4.4.6	实例 26: 生产者用单线程发送消息	107
4.4.7	实例 27: 生产者用多线程发送消息	110
4.5	配置生产者的属性	112
4.6	保存对象的各个属性——序列化	115
4.6.1	实例 28: 序列化一个对象	115
4.6.2	实例 29: 在生产者应用程序中实现序列化	117
4.7	自定义主题分区	122

4.7.1	实例 30: 编写自定义主题分区的算法	122
4.7.2	实例 31: 演示自定义分区类的使用	123
4.8	小结	125
第 5 章	从 Kafka 系统中读取消息数据——消费	126
5.1	本章教学视频说明	126
5.2	了解 Kafka 消费者	126
5.2.1	为什么需要消费者组	126
5.2.1	消费者和消费者组的区别	127
5.2.2	消费者和分区的对应关系	127
5.3	使用 Kafka 系统的脚本操作消费者	130
5.3.1	认识消费者新接口	130
5.3.2	实例 32: 用新接口启动消费者程序, 并查看消费者信息	131
5.3.3	实例 33: 用旧接口启动消费者程序, 并查看消费者元数据的 存储结构	134
5.4	消费 Kafka 集群中的主题消息	136
5.4.1	主题如何自动获取分区和手动分配分区	137
5.4.2	实例 34: 主题自动/手动获取分区	137
5.4.3	实例 35: 反序列化主题消息	140
5.4.4	如何提交消息的偏移量	145
5.4.5	实例 36: 使用多线程消费多个分区的主题	146
5.5	配置消费者的属性	150
5.6	小结	151
第 6 章	存储及管理数据	152
6.1	本章教学视频说明	152
6.2	分区存储数据	152
6.2.1	熟悉分区存储	153
6.2.2	了解消息的格式	154
6.3	清理过期数据的两种方法	155
6.4	网络模型和通信流程	156
6.4.1	基本数据类型	156
6.4.2	通信模型	157

6.4.3 通信过程	157
6.6 小结	159

第3篇 进阶

第7章 Kafka 安全机制	162
----------------------	-----

7.1 本章教学视频说明	162
7.2 了解 Kafka 的安全机制	162
7.2.1 身份验证	163
7.2.2 权限控制	163
7.3 使用 SSL 协议进行加密和身份验证	164
7.3.1 了解 SSL 协议	164
7.3.2 实例 37: 创建 SSL 密钥库, 并查看密钥库文件	165
7.3.3 实例 38: 创建私有证书	167
7.3.4 实例 39: 导出证书, 使用 CA 对证书进行签名	170
7.3.5 实例 40: 在服务端配置 SSL 协议, 并创建主题	173
7.3.6 实例 41: 在客户端配置 SSL 协议, 并读/写数据	174
7.4 使用 SASL 协议进行认证	176
7.4.1 给客户端配置“Java 认证和授权服务”(JAAS)	176
7.4.2 给服务端配置 SASL	178
7.4.3 实例 42: 开启 SASL/Kerberos 认证协议	178
7.4.4 实例 43: 开启 SASL/PLAIN 认证协议	181
7.4.5 实例 44: 开启 SASL/SCRAM 认证协议	184
7.5 权限控制	187
7.5.1 权限控制的基础命令	187
7.5.2 配置 ACL (访问控制列表)	188
7.5.3 实例 45: 启动集群	189
7.5.4 实例 46: 查看授权、添加授权、删除授权	190
7.6 小结	195

第8章 用 Kafka 连接器建立数据管道	196
-----------------------------	-----

8.1 本章教学视频说明	196
8.2 认识 Kafka 连接器	196

8.2.1	了解连接器的使用场景	197
8.2.2	特性及优势	198
8.2.3	连接器的几个核心概念	198
8.3	操作 Kafka 连接器	199
8.3.1	配置 Kafka 连接器的属性	199
8.3.2	认识应用接口——REST API	202
8.3.3	实例 47: 单机模式下, 将数据导入 Kafka 主题中	203
8.3.4	实例 48: 单机模式下, 将 Kafka 主题中的数据导出	205
8.3.5	实例 49: 分布式模式下, 将数据导入 Kafka 主题	206
8.4	实例 50: 开发一个简易的 Kafka 连接器插件	210
8.4.1	编写 Source 连接器	211
8.4.2	编写 Sink 连接器	217
8.4.3	打包与部署	220
8.5	小结	225
第 9 章	Kafka 流处理	226
9.1	本章教学视频说明	226
9.2	初识 Kafka 流处理	227
9.2.1	什么是流处理	227
9.2.2	什么是流式计算	227
9.2.3	为何要使用流处理	228
9.3	了解流处理的架构	229
9.3.1	流分区与任务	230
9.3.2	线程模型	232
9.3.3	本地状态存储	234
9.3.4	容错性 (Failover)	235
9.4	操作 KStream 和 KTable	235
9.4.1	流处理的核心概念	236
9.4.2	窗口操作	237
9.4.3	连接操作	241
9.4.4	转换操作	246
9.4.5	聚合操作	247
9.5	实例 51: 利用流处理开发一个单词统计程序	248

9.5.1	创建 Kafka 流主题	248
9.5.2	统计流主题中单词出现的频率	249
9.5.3	预览操作结果	250
9.6	实例 52：利用 Kafka 流开发一个 SQL 引擎	251
9.6.1	构建生产流数据源	251
9.6.2	构建 Kafka 流处理	253
9.6.3	构建数据结构和执行 SQL 逻辑	254
9.6.4	观察操作结果	255
9.7	小结	256
第 10 章	监控与测试	257
10.1	本章教学视频说明	257
10.2	Kafka 的监控工具——Kafka Eagle 系统	258
10.2.1	实例 53：管理主题	258
10.2.2	实例 54：查看消费者组信息	259
10.2.3	实例 55：查看 Kafka 与 Zookeeper 集群的状态和性能	263
10.3	测试生产者性能	264
10.3.1	了解测试环境	264
10.3.2	认识测试工具	265
10.3.3	实例 56：利用工具测试生产者性能	266
10.4	测试消费者性能	275
10.4.1	了解测试环境	275
10.4.2	认识测试工具	276
10.4.3	实例 57：利用脚本测试消费者的性能	276
10.4	小结	280

第 4 篇 商业实战

第 11 章	Kafka 与 ELK 套件的综合	282
11.1	本章教学视频说明	282
11.2	安装与配置 ELK	283
11.2.1	安装与配置 LogStash	283
11.2.2	实例 58：LogStash 的标准输入与输出	285

11.2.3	安装与配置 Elasticsearch	287
11.2.4	实例 59: 使用 Elasticsearch 集群的 HTTP 接口创建索引	292
11.2.5	实例 60: 使用 Elasticsearch 集群的 HTTP 接口查看索引	293
11.2.6	实例 61: 使用 Elasticsearch 集群的 HTTP 接口添加数据	294
11.2.7	安装与配置 Kibana	296
11.2.8	实例 62: 启动并验证 Kibana 系统	298
11.3	实例 63: 实现一个游戏日志实时分析系统	299
11.3.1	了解系统要实现的功能	300
11.3.2	了解平台体系架构	300
11.3.3	采集数据	302
11.3.4	分流数据	304
11.3.5	实现数据可视化	306
11.4	小结	308
第 12 章	Kafka 与 Spark 实时计算引擎的整合	309
12.1	本章教学视频说明	309
12.2	介绍 Spark 背景	310
12.2.1	Spark SQL——Spark 处理结构化数据的模块	310
12.2.2	Spark Streaming——Spark 核心应用接口的一种扩展	311
12.2.3	MLlib——Spark 的一个机器学习类库	311
12.2.4	GraphX——Spark 的一个图计算框架	311
12.3	准备 Spark 环境	311
12.3.1	下载 Spark 基础安装包	311
12.3.2	安装与配置 Spark 集群	312
12.4	操作 Spark	315
12.4.1	实例 64: 使用 Spark Shell 统计单词出现的频率	315
12.4.2	实例 65: 使用 Spark SQL 对单词权重进行降序输出	317
12.4.3	实例 66: 使用 Spark Submit 统计单词出现的频率	319
12.5	实例 67: 对游戏明细数据做实时统计	322
12.5.1	了解项目背景和价值	323
12.5.2	设计项目实现架构	323
12.5.3	编码步骤一 实现数据采集	325
12.5.4	编码步骤二 实现流计算	327

12.5.5	编码步骤三 打包应用程序.....	330
12.5.6	编码步骤四 创建表结构	332
12.5.7	编码步骤五 执行应用程序.....	332
12.5.8	编码步骤六 预览结果	333
12.6	小结	333
第 13 章	实例 68：从零开始设计一个 Kafka 监控系统——Kafka Eagle	334
13.1	本章教学视频说明	334
13.2	了解 Kafka Eagle 监控系统.....	335
13.2.1	设计的背景	335
13.2.2	应用场景	336
13.3	从结构上了解 Kafka Eagle.....	337
13.3.1	了解 Kafka Eagle 的整体架构和代码结构.....	337
13.3.2	设计 Kafka Eagle 的 7 大功能模块.....	339
13.4	实现 Kafka Eagle 的功能模块.....	347
13.4.1	编码步骤一 实现数据面板.....	347
13.4.2	编码步骤二 实现主题管理.....	348
13.4.3	编码步骤三 实现消费者实例详情.....	350
13.4.4	编码步骤四 实现集群监控.....	350
13.4.5	编码步骤五 实现性能监控.....	351
13.4.6	编码步骤六 实现告警功能.....	351
13.4.7	编码步骤七 实现系统功能.....	352
13.5	安装及使用 Kafka Eagle 监控系统.....	353
13.5.1	准备环境	353
13.5.2	快速部署	354
13.5.3	了解 Kafka Eagle 的基础命令.....	358
13.6	小结	358

的“多生产者多消费者”（Too Many Connections）问题。

因此，消息队列系统应该提供一个生产者模型，它允许多个生产者同时向一个消息队列发送消息，并且该消息队列能够处理这些消息。

第1章

1.2.1 什么是消息队列

第1篇 准备

业界已有的消息队列系统，在扩展性、可靠性、资源利用率和吞吐率方面存在明显不足，故 LinkedIn 团队开始尝试设计一款新的消息队列系统，进而诞生了 Kafka 消息队列系统。

本篇将介绍学习 Kafka 前的准备工作，包括 Kafka 的基本概念和架构，以及详细的安装步骤。

► 第1章 了解消息队列和 Kafka

► 第2章 安装及配置 Kafka

• 消息队列，即比生产者稍后处理，并保证消息不丢失。

• 消息队列，即比生产者稍后处理，并保证消息不丢失。

• 消息队列，即比生产者稍后处理，并保证消息不丢失。

• 消息队列，即比生产者稍后处理，并保证消息不丢失。

1. 应用解耦

由于消息队列的存在，

应用可以独立开发，

例如，用户信息

（1）用户信息

（2）用户信息



生产者

消息队列

消费者

生产者

消息队列

消费者

生产者

消息队列

• 生产者，即比生产者稍后处理，并保证消息不丢失。

• 生产者，即比生产者稍后处理，并保证消息不丢失。

第 1 章

了解消息队列和Kafka

本章的知识都是 Kafka 基础,学习起来会非常轻松。本章能够帮助读者从零开始认识 Kafka,内容包含消息队列、Kafka 的起源、Kafka 的基础知识等。

1.1 本章教学视频说明

视频内容: 什么是消息队列、消息队列与 Kafka 之间的联系、Kafka 的基本概念、Kafka 的工作机制,以及 Kafka 的使用范围等。

视频时长: 10 分钟。

视频截图见图 1-1。

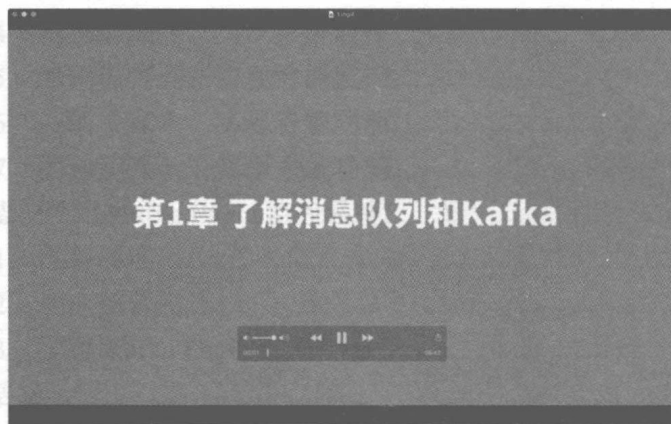


图 1-1 本章教学视频截图

1.2 消息队列

在高并发的应用场景中,由于来不及同步处理请求,接收到的请求往往会发生阻塞。例如,大量的插入、更新请求同时到达数据库,这会导致行或表被锁住,最后会因为请求堆积过多而